AA2

clones, defining an ORF of [2842] 2843 amino acids. In frame stop codons surrounded this ORF, as described in the text, suggesting that the entire APC gene product was represented in the ORF illustrated. Only the predicted amino acids are shown.

At column 6, line 30:

AA3

Alteration of wild-type genes can also be detected on the basis of the alteration of a wild-type expression product of the gene. Such expression products include both the APC mRNA as well as the APC protein product. The sequences of these products are shown in [FIG. 3] FIGS. 3A-3Z. Point mutations may be detected by amplifying and sequencing the mRNA or via molecular cloning of cDNA made from the mRNA. The sequence of the cloned cDNA can be determined using DNA sequencing techniques which are well known in the art. The cDNA can also be sequenced via the polymerase chain reaction (PCR) which will be discussed in more detail below.

At column 8, line 32:

AA4

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from APC sequences or sequences adjacent to APC except the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the APC open reading frame shown in [FIG. 3] FIGS. 3A-3Z (SEQ ID NO: 1), design of particular primers is well within the skill of the art.

At column 10, line 39:

Polypeptides which have APC activity can be supplied to cells which carry mutant or missing APC alleles. The sequence of the APC protein is disclosed in [FIG. 3] FIGS. 3A-3Z (SEQ ID NO:7). [These two sequences differ slightly and appear to be indicate the existence of two different forms of the APC protein.] Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, APC can be extracted from APC-producing mammalian cells such as brain cells. In addition, the techniques of synthetic chemistry can be employed to synthesize APC protein. Any of such techniques can provide the preparation of the present invention which comprises the APC protein. The preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or *in vitro*.

AA5

At column 10, line 66:

A short region of homology has been identified between APC and the human m3 muscarinic acetylcholine receptor (mAChR). This homology was largely confined to 29 residues in which 6 out of 7 amino acids (EL(GorA)GLQA) were identical (See [FIG. 4] FIG. 4B (SEQ ID NO: 9)). Initially, it was not known whether this homology was significant, because many other proteins had higher levels of global homology (though few had six out of seven contiguous amino acids in common). However, a study on the sequence elements controlling G protein activation by mAChR subtypes (Lechleiter et al., EMBO J., p. 4381 (1990)) has shown that a 21 amino acid region from the m3 mAChR completely mediated G protein specificity when substituted for the 21 amino acids of m2 mAChR at the analogous protein position. These 21 residues overlap the 19 amino acid homology between APC and m3 mAChR.

AA6

At column 13, line 1:

Del 9"

Contig 2: TB1 - TB1 was identified through a cross-hybridization approach. Exons of genes are often evolutionarily conserved while introns and intergenic regions are much less conserved. Thus, if a human probe cross-hybridizes strongly to the DNA from non-primate species, there is a reasonable chance that it contains exon sequences. Subclones of the cosmids shown in [FIG. 1] FIGS. 1A, 1B-1, 1B-2, and 1B-3 were used to screen Southern blots containing rodent DNA samples. A subclone of cosmid N5.66 (p 5.66-4) was shown to strongly hybridize to rodent DNA, and this clone was used to screen cDNA libraries derived from normal adult colon and fetal liver. The ends of the initial cDNA clones obtained in this screen were then used to extend the cDNA sequence. Eventually, 11 cDNA clones were isolated, covering 2314 bp. The gene detected by these clones was named TB1. Sequence analysis of the overlapping clones revealed an open reading frame (ORF) that extended for 1302 bp starting from the most 5' sequence data obtained (FIG. 2A). If this entire open reading frame were translated, it would encode 434 amino acids (SEQ ID NO: 5). The product of this gene was not globally homologous to any other sequence in the current database but showed two significant local similarities to a family of ADP, ATP carrier/translocator proteins and mitochondrial brown fat uncoupling proteins which are widely distributed from yeast to mammals. These conserved regions of TB1 (underlined in FIG. 2A) may define a predictive motif for this sequence family. In addition, TB1 appeared to contain a signal peptide (or mitochondrial targeting sequence) as well as at least 7 transmembrane domains.

AA7

At column 14, line 38:

Sequence analysis of the APC cDNA clones revealed an open reading frame of 8,535 nucleotides. The 5' end of the ORF contained a methionine codon (codon 1) that was preceded

AA8

by an in-frame stop codon 9 bp upstream, and the 3' end was followed by several in-frame stop codons. The protein produced by initiation at codon 1 would contain [2,842] <u>2843</u> amino acids (SEQ ID NO: 7) [(FIG. 3)] <u>FIG. 3A-3Z</u>. The results of database searching with the APC gene product were quite complex due to the presence of large segments with locally biased amino acid compositions. In spite of this, APC could be roughly divided into two domains. The N-terminal 25% of the protein had a high content of leucine residues (12%) and showed local sequence similarities to myosins, various intermediate filament proteins (e.g., desmin, vimentin, neurofilaments) and Drosophila armadillo/human plakoglobin. The latter protein is a component of adhesive junctions (desmosomes) joining epithelial cells (Franke et al., Proc. Natl. Acad. Sci. U.S.A., Vol. 86, p. 4027 (1989); Perfer et al., Cell, Vol. 63, p. 1167 (1990)). The C-terminal 75% of APC (residues 731-2832) is 17% serine by composition with serine residues more or less uniformly distributed. This large domain also contains local concentrations of charged (mostly acidic) and proline residues. There was no indication of potential signal peptides, transmembrane regions, or nuclear targeting signals in APC, suggesting a cytoplasmic localization.

AA8

At column 26, line 27:

To obtain DNA sequence adjacent to the exons of the genes DP1, DP2.5, and SRP19, sequencing substrate was obtained by inverse PCR amplification of DNAs from two YACs, 310D8 and 183H12, that span the deletions. Ligation at low concentration cyclized the restriction enzyme-digested YAC DNAs. Oligonucleotides with sequencing tails, designed in inverse orientation at intervals along the cDNAs, primed PCR amplification from the cyclized templates. Comparison of these DNA sequences with the cDNA sequences placed exon boundaries at the divergence points. SRP19 and DP1 were each shown to have five exons. DP2.5 consisted of 15

AA9

exons. The sequences of the oligonucleotides synthesized to provide PCR amplification primers for the exons of each of these genes are listed in Table III [SEQ ID NOS:39-94] <u>(SEQ ID NOS: 39-94)</u>. With the exception of exons 1, 3, 4, 9, and 15 of DP2.5 (see below), the primer sequences were located in intron sequences flanking the exons. The 5' primer of exon 1 is complementary to the cDNA sequence, but extends just into the 5' Kozak consensus sequence for the initiator methionine, allowing a survey of the translated sequences. The 5' primer of exon 3 is actually in the 5' coding sequences of this exon, as three separate intronic primers simply would not amplify. The 5' primer of exon 4 just overlaps the 5' end of this exon, and we thus fail to survey the 19 most 5' bases of this exon. For exon 9, two overlapping primer sets were used, such that each had one end within the exon. For exon 15, the large 3' exon of DP2.5, overlapping primer pairs were placed along the length of the exon; each pair amplified a product of 250-400 bases.

AA9

At column 29, line 1:

The sequences of the unique conformers from exons 7, 8, 10, and 11 of DP2.5 revealed dramatic mutations in the DP2.5 gene. The sequence of the new mutation creating the exon 7 conformer in patient 3746 was shown to contain a deletion of two adjacent nucleotides, at positions 730 and 731 in the cDNA sequence ([FIG. 7,] SEQ ID NO:1). The normal sequence at this splice junction is <u>CAG</u>GGTCA (intronic sequence underlined), with the intron-exon boundary between the two repetitions of AG. The mutant allele in this patient has the sequence CAGGTCA. Although this change is at the 5' splice site, comparison with known consensus sequences of splice junctions would suggest that a functional splice junction is maintained. If this new splice junction were functional, the mutation would introduce a frameshift that creates a stop

AA10

codon 15 nucleotides downstream. If the new splice junction were not functional, messenger processing would be significantly altered.

At column 29, line 26:

The unique conformer found in exon 8 of patient 3460 was found to carry a C-T transition, at position 904 in the cDNA sequence of DP2.5 [(shown in FIG. 7)], which replaced the normal sequence of CGA with TGA. This point mutation, when read in frame, results in a stop codon replacing the normal arginine codon. This single-base change had occurred within the context of a CG dimer, a potential hot spot for mutation (Barker et al., 1984).

AA11

At column 30, line 37:

The continuity of the very large (6.5 kb), most 3' exon in DP2.5 was shown in two ways. First, inverse PCR with primers spanning the entire length of this exon revealed no divergence of the cDNA sequence from the genomic sequence. Second, PCR amplification with converging primers placed at intervals along the exon generated products of the same size whether amplified from the originally isolated cDNA, cDNA from various tissues, or genomic template. Two forms of exon 9 were found in DP2.5: one is the complete exon; and the other, labeled exon 9A, is the result of a splice into the interior of the exon that deletes bases 934 to 1236 in the mRNA and removes 101 amino acids from the predicted protein (see [FIG. 3] FIGS. 3A-3Z, SEQ ID NOS: 1 & 2).

AA12

At column 31, line 30:

The cDNA consensus sequence of APC predicts that the longer, more abundant form of the message codes for a [2842 or 2844] 2843 amino acid peptide with a mass of 311.8 kd. This predicted APC peptide was compared with the current data bases of protein and DNA sequences using both Intelligenetics and GCG software packages. No genes with a high degree of amino

AA13

acid sequence similarity were found. Although many short (approximately 20 amino acid) regions of sequence similarity were uncovered, none was sufficiently strong to reveal which, if any, might represent functional homology. Interestingly, multiple similarities to myosins and keratins did appear. The APC gene also was scanned for sequence motifs of known function; although multiple glycosylation, phosphorylation, and myristoylation sites were seen, their significance is uncertain.

AA13

At columns 31-132:

Please delete the sequence listing and replace it with the enclosed substitute sequence listing. The substitute sequence listing is identical to the sequence listing in the patent with the exception of one amino acid in SEQ ID NO:7. The substitute sequence listing contains a proline at position 173.

Remarks

The specification has been amended to correct the number of amino acids said to be present in the APC protein. This correction is supported in Figure 3 and in SEQ ID NOS:1 and 2, each of which show a 2843 amino acid APC protein.

The sequence listing has been amended to correct the amino acid sequence of the APC protein shown in SEQ ID NO:7, by insertion of a proline at position 173 of SEQ ID NO:7. This amendment is supported in the issued patent in Figure 3 and in SEQ ID NOS:1 and 2, each of which contain a proline at position 173. A computer readable form of the substitute sequence listing is provided for use in examining this application. The contents of the computer readable form and the paper copy of the substitute sequence listing are identical. The contents of the substitute sequence listing are identical to those of the original sequence listing except for the insertion of the proline at position 173